The Influence of Numerical Anchors on Estimation Accuracy: An Experimental Approach

Group 2 - Ahmed Farid, Luca Matteucci, Santiago Mazzei, Srithijaa Sankepally, Victor Floriano

03/07/2024

1. Abstract

This study investigates the anchoring effect, a cognitive bias where initial information influences individuals' estimates, regardless of its relevance. We particularly investigate how high and low numerical anchors impact individuals' accuracy in estimating the quantity of items in a container. By conducting a randomized experiment with varied anchor points, we assess the effect of these anchors on estimation. Our results enhance the comprehension of cognitive biases in decision-making and estimation processes.

2. Introduction

The anchoring effect is a central concept in cognitive psychology, indicating that an initial piece of information, or "anchor," can profoundly influence subsequent judgments and decisions. This study is driven by the hypothesis that numerical anchors, whether high or low, affect individuals' accuracy in unrelated estimation tasks. Investigating this phenomenon is essential to understand the mechanisms of human judgment and decision-making. Our null hypothesis asserts no significant difference in mean estimates between participants exposed to high versus low numerical anchors. Conversely, our alternative hypothesis suggests a significant difference, indicating that numeric anchors impact participants' estimates.

3. Literature Review

Previous research on the anchoring effect highlights its significant influence across various domains, from financial decision-making to social judgments. *Tversky and Kahneman (1974)* first identified this cognitive bias, demonstrating its pervasive impact on decision-making. Subsequent studies have explored the effect in different contexts. *"Implications of Attitude Change Theories for Numerical Anchoring: Anchor Plausibility and the Limits of Anchor Effectiveness"(Wegener et al.)*, explores how extreme versus moderate numerical anchors affect judgments. Contrary to traditional beliefs that extreme anchors lead to stronger effects, this study finds that moderate anchors can have a more significant impact. This is aligned with attitude change theories, suggesting that the plausibility of an anchor influences its effectiveness. The findings challenge conventional "anchor-and-adjust" and "selective accessibility" views, indicating a need for a more nuanced understanding of anchoring effects in decision-making processes.

Another study, *"Subliminal Anchoring: Judgmental Consequences and Underlying Mechanisms" (Mussweiler and Englich)*, explores the concept of subliminal anchoring, where judgment is influenced by numeric standards presented outside of conscious awareness. Through various experiments, the authors demonstrate that such subliminal anchors can sway people's estimates and judgments in a similar manner to explicit anchors. The research underscores the power of subconscious stimuli in shaping decision-making processes and emphasizes the role of subliminal cues in guiding human judgment without their explicit recognition.

The insights from our literature review not only deepened our understanding of the topic but also enabled us to utilize the data from these studies to estimate an expected Cohen's D for our experiment, which will be elaborated on later.

4. Methodology and Experimental Design

Our survey (Appendix 1) was distributed online, leveraging a wide range of platforms to ensure a diverse and extensive reach. It was shared across various WhatsApp groups, social media channels, Reddit forums, and other digital platforms to attract participants from different demographics and backgrounds. This strategic distribution was aimed at garnering a broad spectrum of responses, enriching the data pool for more comprehensive analysis.

Survey participants are first informed about the fictional Lars Halden's skill in estimating item quantities. They are then randomly divided, through simple randomization, into three groups to study anchoring's psychological effects. The high anchor group learns of Lars' feats at age 102, setting a high benchmark, while the low anchor group hears about his abilities at age 9, establishing a lower standard. The control group gets neutral details about Lars' Norwegian background, without age anchors. Following the introduction, participants are engaged with questions related to the story to ensure they have absorbed the details and to enhance their engagement with the survey process. Specifically, those in the high and low anchor groups are asked about Lars' age, reinforcing the anchoring effect, whereas control group participants are queried about his name to maintain consistency in participant engagement across all groups. These questions were also used as an attention check for our survey.

The core of the survey involves a series of 6 estimation tasks, where participants view images of containers filled with various items, including basketballs, M&Ms, coins, sequence chips, puzzle pieces, and pistachios. Each participant is tasked with estimating the quantity of items in each container. To assist in this process, a reference item is placed outside each container, providing a tangible comparison point to aid in their estimations. The presentation order of the images is standardized across all participants, beginning with the basketball in a rim image. This deliberate order is based on the assumption that starting with a seemingly attainable estimation task may encourage participants to apply a similar level of accuracy to subsequent tasks.

Finally, the survey concludes with optional demographic questions about the participants' gender and age.

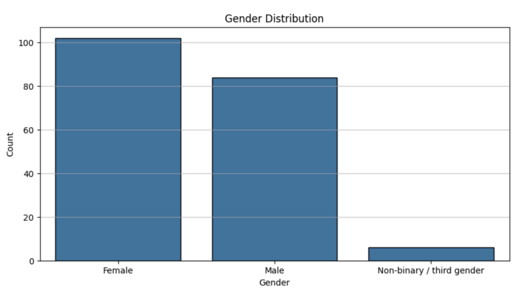
5. Data Collection/Structure

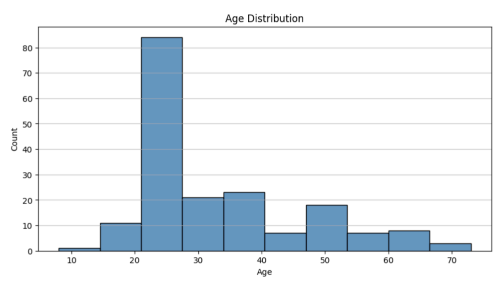
Our initial power analysis determined that each group needed at least 59 participants. This calculation was based on an average Cohen's D of 0.79 from four previous studies on the anchoring effect, aiming for a statistical power of 95% and a significance level of 1%.

Initially, 219 participants were enrolled and completed the study, but after identifying and removing duplicate entries based on IP addresses - retaining only the first entry per duplicate - the total participant count was adjusted to 192. This measure was crucial for maintaining the uniqueness of each participant and preserving the integrity of the collected data. The distribution of participants across the study groups was as follows: The control group comprised 70 participants. The first treatment group, referred to as the low anchor group, included 49 participants. The second treatment group, known as the high anchor group, had 73 participants.

Out of 192 participants, 83 responded to the demographic questions. Here are some of the reported results:

- The average age of the participants is 32.63 years, with a standard deviation of 13.20 years. However, the median age is 27 years, indicating a relatively young sample.
- Regarding gender distribution, females constitute the majority of the participants, although the distribution across genders is fairly balanced.





6. Data Analysis

6.1 Pre-Experiment Randomization Check

Upon analyzing the data from our survey, we conducted two statistical tests to assess the effectiveness of the randomization. The first test, a proportions z-test, was applied to compare the Control group with both Treatment groups. The obtained p-value was 0.37, leading us to conclude that the randomization was adequately conducted, as we could not reject the null hypothesis of proper randomization.

In the second analysis, we conducted regression tests, evaluating the covariates Gender (1) and Age (2) across different treatment arms. In the figure below, Gender (1) regression indicated no significant coefficients for either the Low or High Anchor groups, affirming successful randomization concerning gender. However, in the Age (2) regression, the Low Anchor group showed a significant coefficient (-5.2), suggesting these participants were, on average, younger than those in the Control group.

	Gender Model (1)	Age Model (2)
Low Anchor	-0.030	-5.261^{**}
	(0.095)	(2.485)
High Anchor	0.009	-3.404
	(0.085)	(2.256)
Intercept	0.456^{***}	35.303^{***}
	(0.061)	(1.613)
Observations	186	183
Note:	*p<0.1; **p<0.05; ***p<0.01	

Further, we conducted the same regression analyses on a subset of data, specifically involving participants from the treatment arms (Appendix 2). These analyses showed no significant coefficients for either covariate (Gender or Age) between the Low and High Anchor groups, indicating effective randomization within the treatment arms.

While the Age (2) regression suggests a variation in average age for the Low Anchor group compared to Control, the overall data supports the success of the experiment's randomization process.

6.2 Data Structure Challenge

Before analyzing the Average Treatment Effect (ATE) for each treatment group, we faced a data structure issue. Our survey collected six different estimates from each participant, one for each estimation task (image). To adapt this for regression analysis, where we required a single estimate as the outcome variable, we transformed our original dataset. Originally, it contained multiple estimates per participant (identified by 'IPAddress') in one row; we converted this into a format with single estimates per row, as seen below:

	IPAddress	Gender	Age	group	Question	Score
0	77.164.41.88	Female	22.0	treatment 1	Q1	11

A challenge with this transformation is the lack of independence among estimates, as each participant generated six estimates, violating the independence assumption of Ordinary Least Squares (OLS) regression. To address this, we applied clustered standard errors in all our regression models, clustering by the 'IPAddress' variable to recognize unique participants.

6.3 Regression Analysis

To calculate the ATE (average treatment effect) of low and high anchor groups ran two main regressions. As discussed, the regressions used clustered standard errors to account for participants making multiple guesses. Additionally, we opted to use the natural logarithm of the estimate values instead of raw estimates, anticipating a multiplicative treatment effect. Recognizing that different images (questions) could influence baseline estimates (e.g., a basketball figure versus an M&Ms jar), we also incorporated "Image" as a fixed effect in our regression models to control for these variations.

Regression 1: Log(Estimate) on Treatment

In the regression below we regressed the log of estimates on treatment, including all the participants from the survey. From the regression output we got an ATE of -3.1% (approximately) for the Low Anchor group, and an ATE of -18.1% for the High Anchor group. That meant that on average, the Low Anchor group had estimates below that on the Control, the same was true for the High Anchor group. While this is not what we would have expected, as the initial expectation was that a high anchor would lead to higher estimates, we can see that both coefficients are not statistically significant. Therefore, we actually have no evidence that average estimates in either treatment group are different from that of the control group.

DEPVAR	LOG(ESTIMATE)
LOW ANCHOR	-0.031 (0.100)
HIGH ANCHOR	-0.181 (0.094)
IMAGE	X
R2	0.668
S.E. TYPE	BY: IPADDRESS
OBSERVATIONS	1151

SIGNIFICANCE LEVELS: * P < 0.05, ** P < 0.01, *** P < 0.001
FORMAT OF COEFFICIENT CELL: COEFFICIENT (STD. ERROR)

Regression 2: Log(Estimate) on Treatment - participants that passed the attention check

For our second regression, we filtered our data to include only those participants who passed our attention check question. That was the question asking them to recall the age (treatment arms), or name (control) of the character described in the story. On this iteration we got different ATEs, Low Anchor with +1.7% and High Anchor with -8.5% compared to Control. Similarly to "Regression 1" the results were not what we expected, but again, neither coefficient was statistically significant.

DEPVAR	LOG(ESTIMATE)
LOW ANCHOR.	0.017 (0.065)
HIGH ANCHOR	-0.085 (0.059)
IMAGE	X
R2	0.566
S.E. TYPE	BY: IPADDRESS
OBSERVATIONS	1042

SIGNIFICANCE LEVELS: * P < 0.05, ** P < 0.01, *** P < 0.001
FORMAT OF COEFFICIENT CELL: COEFFICIENT (STD. ERROR)

Regression 3: Log(Estimate) on Low vs. High Anchor

For our third and final regression we selected only those participants who had been assigned to either the Low Anchor group or the High Anchor group to focus on the difference between estimates for low and high anchors. From the output below, the ATE of seeing a high anchor, compared to low, was -20.7%. On the same line as the previous to regression this went against our initial expectations, but the coefficient was not statistically significant, indicating that there is no evidence that Low Anchor and High Anchor groups had any significant difference on their average estimates.

```
            ------------------------------------
            DEPVAR                 LOG(ESTIMATE)
            ------------------------------------
            HIGH ANCHOR (VS LOW.) -0.207 (0.106)
            ------------------------------------
            IMAGE                          X
            ------------------------------------
            R2                             0.670
            S.E. TYPE              BY: IPADDRESS
            OBSERVATIONS                     654
            ------------------------------------
SIGNIFICANCE LEVELS: * P < 0.05, ** P < 0.01, *** P < 0.001
                FORMAT OF COEFFICIENT CELL:
                  COEFFICIENT (STD. ERROR)
```

Finally, the results from regressions 1 to 3 presented no evidence that exposure to low or high anchors changed the average estimate of survey participants compared to those in control (no anchor), or between the two treatment groups. We failed to reject our null hypothesis that the numeric anchors have no effect on mean estimates.

6.4 The Power of Regression 3

Using the mean scores and standard deviations of the low anchor and high anchor groups, we were able to calculate Cohen's D value of 0.11 for our "Regression 3". After calculating Cohen's D, we were able to calculate the power of our experiment at 22%. The low values of both Cohen's D and power can be explained by the high variance in the scores of respondents, which will be discussed further in the next sections.

7. Discussion and Limitations

The findings from our experimental analysis are inconclusive in proving that numerical anchors affect people's accuracy in pictorial estimation tasks. These results were somewhat surprising to us given the substantial effect sizes we had calculated using previous research. We calculated that we would need 59 people per treatment arm (177 total sample size) to detect an effect, if such an effect existed. Given that we were able to gather responses from 192 participants, we would expect to detect an effect. However, this was not the case and we believe there are two main reasons for this.

1. Variance: There was a large degree of variance amongst the responses given by the participants. For the image with coins in the glass, most of us agreed that a 'reasonable estimate' would be somewhere between 50 and 120, however we had many responses

which showed little thought such as 1000 and 2000. Due to these outlier responses (in both the treatment group and the control group), we can not reasonably assume that people are being affected by the anchor much, if at all.

2. Interference: The way the survey was distributed leaves room for interference in the responses. Most survey responses are from people who were sent the survey in groups. This was done to get the most responses possible by creating competition amongst friend groups. However, this increases the possibility of discussions amongst the group and thereby is likely to reduce the difference between the estimates of the treatment arms.

There are a few solutions to account for these problems and increase the likelihood of getting a significant result, in line with previous research. The first solution is to conduct the survey in person. This will make sure people take the survey more seriously, which will reduce the variance amongst the survey respondents' answers. Also, it will help reduce interference as the conductors of the experiment will have greater control over the interactions between participants of the survey. Apart from conducting the survey in person, offering the respondents an incentive to complete the survey is also likely to increase the investment they feel while doing the survey and, as a result, most likely reduce the variance of the responses.

To conclude, this experiment lays a basis for conducting anchoring effect studies on pictorial estimations in the future. It underlines the importance of respondent investment in the survey as well as the need to adopt more robust methodologies to find effects that likely exist as evidenced in the prior literature.

References

Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases. Science, 185(4157), 1124-1131.

Mussweiler, Thomas, and Birte Englich. "Subliminal Anchoring: Judgmental Consequences and Underlying Mechanisms." Organizational Behavior and Human Decision Processes, Academic Press, 12 Sept. 2005, www.sciencedirect.com/science/article/pii/S0749597805001093?ref=pdf_download&fr=RR-2&rr=860d9459dc6f6aca.

Wegener, Duane T., et al. Implications of Attitude Change Theories for Numerical ..., 15 Mar. 2000, richardepetty.com/wp-content/uploads/2019/01/2001-jesp-wegenerpettybedelljarvis.pdf.

Appendix

Appendix 1 - Survey

BOSTON UNIVERSITY

Story

Norwegian-born **Lars Halden** demonstrated unparalleled skill in competitions for container item estimation, achieving international recognition. His success led to the initiation of a study to investigate whether such precise estimating skills are innate or can be developed, highlighting the universal potential for excellence.

Now, we're **conducting a survey** to understand how well the general population can estimate items in a container, inspired by Lars' legendary precision. Your participation will help us uncover if Lars' gift is truly unique or if it's a skill that many can develop with practice.

At just **9 years old**, **Lars Halden** demonstrated unparalleled skill in competitions for container item estimation, illustrating extraordinary abilities in his youth. His success led to the initiation of a study to investigate whether such precise estimating skills are innate or can be developed, highlighting the potential at an early stage of life.

Now, we're **conducting a survey** to understand how well the general population can estimate items in a container, inspired by Lars' legendary precision. Your participation will help us uncover if Lars's gift is truly unique or if it's a skill that many can develop with practice.

At the **age of 102**, **Lars Halden** demonstrated unparalleled skill in competitions for container item estimation, illustrating that extraordinary abilities are not limited by age. His success led to the initiation of a study to investigate whether such precise

estimating skills are innate or can be developed, highlighting the potential in every stage of life.

Now, we're **conducting a survey** to understand how well the general population can estimate items in a container, inspired by Lars' legendary precision. Your participation will help us uncover if Lars' gift is truly unique or if it's a skill that many can develop with practice.

Block 8

How old is the person in the story? (feel free to go back if you don't remember)

How old is the person in the story? (feel free to go back if you

don't remember)

What is the name of the person in the story? (feel free to go back if you don't remember)

Instruction

Now you'll be asked to estimate how many items are in different containers. Please enter your best guess in the designated box below each image

Default Question Block

Basketballs in a rim

Estimate how many basketballs are in the rim.

Default Question Block

M&Ms in a jug

Estimate how many M&M's are in the jug (there is one M&M

outside the jug for reference).

Default Question Block

Coins in a glass

Estimate how many coins are in the glass (there is one coin

outside the glass for reference).

Default Question Block

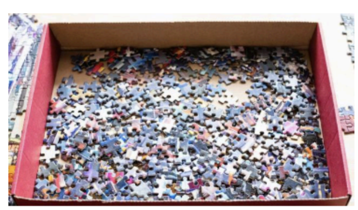
Sequence chips in a glass

Estimate how many chips are in the glass (there is one chip

outside the glass for reference).

Default Question Block

Puzzle pieces inside the box



07/03/2024, 21:19 Qualtrics Survey Software

Estimate how many puzzle pieces are INSIDE the box.

Default Question Block

Pistachios in a jar

https://bostonu.qualtrics.com/Q/EditSection/Blocks/Ajax/GetSurveyPrintPreview?ContextSurveyID=SV_8wT7eV87fJWPOZw&ContextLibraryID=UR_3HJCOlhEZyPcfSS 13/15

07/03/2024, 21:19 Qualtrics Survey Software

https://bostonu.qualtrics.com/Q/EditSection/Blocks/Ajax/GetSurveyPrintPreview?ContextSurveyID=SV_8wT7eV87fJWPOZw&ContextLibraryID=UR_3HJCOlhEZyPcfSS 14/15

07/03/2024, 21:19 Qualtrics Survey Software

Estimate how many pistachios are in this jar.

Block 8

What gender do you identify with?

- Male
- Female
- Non-binary / third gender

Enter your age

Powered by Qualtrics

https://bostonu.qualtrics.com/Q/EditSection/Blocks/Ajax/GetSurveyPrintPreview?ContextSurveyID=SV_8wT7eV87fJWPOZw&ContextLibraryID=UR_3HJCOlhEZyPcfSS 15/15

Appendix 2 - Randomization Check on Genre and Age Covariates for Participants on Low or High Anchor Groups

	Gender Model (1)	Age Model (2)
Treatment 2	0.039	1.857
	(0.094)	(2.402)
Treatment 1	0.426***	30.042***
	(0.073)	(1.845)
Observations	118	117
Note:	*p<0.1; **p<0.05; ***p<0.01	